Malaysian Journal of Library & Information Science, Vol. 17, no. 2, August 2012: 83-100

Scientometric evaluation of Sankhyá – the Indian Journal of Statistics

Prabir Kumar Das and **Jiban K. Pal**
Library, Documentation & Information Science Division,
Indian Statistical Institute, 203, B. T. Road,
Kolkata – 700108, INDIA
e-mail: prabirdas2003@yahoo.com; jiban@isical.ac.in

ABSTRACT

This paper critically analyses 199 peer-reviewed articles published in Sankhyā during 2003 to 2007. It examines authorship pattern, collaboration trend among authors, predominant areas of statistical research, and time lag in publications. Subsequent analysis focuses on prolific contributors, degree of collaboration, collaboration density, active sub-domains of statistics and time lag trend. Findings reveal the following: (a) the number of articles reduced from 24.6% to 14.0% that conforms to the growth trend of statistical publications in India; (b) single-authored paper counts only 30%, the rest in collaboration either by two-authors (47%) or three-to-five-authors (23%) and average authorship accounts for 1.96 per paper; (c) contributors of Sankhyā worked in highly collaborative manner and the degree of collaboration (CC=0.698) is quite significant; and (d) most of the bilateral and multilateral collaborations has emanated from 12 institutions of 5 different countries. Ranked list of prolific authors has been carried out using fractional counting method. It is observed that author productivity is not in agreement with Lotka's law, but productivity distribution data partially fits the law when the value of α approximated to 2.77 and the number of papers does not exceed two. Broad subject clusters, such as statistics (153) and probability theory (38) constituted about 96% of the contributed articles. Nonparametric inference (18%), parametric inference (15%), design of experiments (10%) and multivariate analysis (8%) are found to be active areas of research in statistics. The study shows an average time lag of fifteen months to publish an article, and a declining trend of time lags following second-degree polynomial type has been observed in this scholarly journal.

Keywords: Single journal studies; Scientometrics analysis; Statistics journal; Sankhyā-the Indian Journal of Statistics; Lotka's law.

INTRODUCTION

Various disciplines have the tradition in measuring research output and intellectual influence of their research community through the studies involving publication productivity and impact using bibliometric analysis. However, researchers as well as research institutions are increasingly evaluated based on their publications produced in the peer-reviewed journals of a particular discipline. In fact the measurement of statistical research becomes a crucial issue over other disciplines, as statistics is the universal tool of inductive inference and technological applications. Different bibliometric methods are extensively used in measuring research outputs and to study the behaviour of scientific disciplines or scientometrics. However, mapping of literature of a particular discipline over a period of time depicts the changes in the cognitive structure and composition of that discipline. Even such studies are essentially predominant for evaluating research activities

and scientific productivity as well as nurturing scientific information. Of late, single-journal bibliometric studies are more prevalent among various levels of research communications, since a journal is considered as sample representative of all scientific communications in a particular discipline. In fact a number of single journal studies have been carried out in multiple dimensions to distinguish the authorship patterns, trends in collaborations, prolific areas of research (Anyi, Zainab and Anuar 2009). Reportedly, such studies are rarely made in the field of Statistics. An attempt has been made to evaluate 'Sankhyā', the first Indian Journal of Statistics with international recognition, aiming in view to map the research on statistical science quantitatively using bibliometric methods and techniques.

SCOPE AND OBJECTIVES

This study is confined to the publications appeared in Sankhyā during the period 2003 to 2007. The study is conducted purely based on journal articles on the major areas of statistics including mathematical statistics and probability. Therefore the less scholarly communications such as book reviews, preface, editorial notes, letters to editor, corrigendum and obituary are excluded from the purview of this study. It is indeed essential to mention that the journal in its' seventy-five years of journey has undergone several changes; by means of splitting into a number of series (A, B, C, and D), and gradually it squeezed into a single-title as appeared in 2003. Further splitting was made in 2008 into two different series with varying scope and ISSN, as decided by the Council of the Institute. Series-A (theoretical statistics) primarily covers the developments in the area of probability theory, stochastic process, and statistical inferences. Series-B (applied statistics) encompasses interdisciplinary research including genomics, bio-informatics, clinical trials, sociometry, biometry, econometrics, demography, sample surveys, statistical computing & data mining and operations research. In fact, in its metamorphic phases, this journal has created significant queries among the statisticians and bibliometricians as well. Therefore the efforts have been pursued to catch a glimpse of Sankhya with the 'unified title' during 2003 to 2007. Hence the bibliometric analysis of peer-reviewed research articles appeared in five volumes (65 to 69, covering twenty issues) of Sankhyā would certainly be an indicative of current trends of statistical research.

The study is intended to investigate the recent trends in statistical research for enabling good research governance by the stakeholders and research scientists in this field. The objectives of this study are as follows:

a) To enumerate the chronological distribution of contributions and to predict the growth trend of statistical publications.
b) To examine the authorship pattern and degree of collaboration in statistical science research.
c) To prepare a ranked list of prolific authors within the dataset studied here and to test the applicability of Lotka's law for author productivity.
d) To determine the extent of collaborative research among the authors, countries, and institutions of statistical science research.
e) To analyze the scattering of publications in broad subject clusters and to detect the active sub-domains of research in statistics.
f) To identify the time lag for publishing an article and average time lags occurred in each publication of this scholarly journal.
g) To justify the importance of this source journal on the growth of statistical research and to enumerate various issues quantitatively relating to the study.

MATERIALS AND METHOD

Sankhyā – The Indian Journal of Statistics (ISSN: 0972-7671), is an international scholarly journal initially published by the Statistical Publishing Society as an organ of Indian Statistical Institute (an institute of national importance). Thereafter this peer-reviewed journal is published by the Indian Statistical Institute, however, Springer Verlag the German giant publishing company has signed an agreement (in 2009) to co-publish *Sankhyā* allowing researchers to access through a global platform Springer-Link. It is the first Indian journal on statistics founded in June 1933 with the editorship of Prasanta Chandra Mahalanobis, (1893-1972), an Indian scientist and applied statistician, who is best remembered for the Mahalanobis distance, a statistical measure. Professor Mahalanobis was the editor-in-chief of this journal until his demise in 1972. The journal emerged to extend the unique perception of the "Professor" toward consideration of statistics as key-technology and to unfold the twin aspects of statistics, both theoretical and applied (Rudra 1996). However it carried much of the path-breaking research works of P. C. Mahalanobis and his close associates such as R. C. Bose, S. N. Roy, S. N. Bose, and C. R. Rao. In pursuance of this philosophy, the journal provides an excellent communication channel for exchanging innovative ideas and developments in different dimensions of statistics, which make Sankhyā an effective and reliable representation of current statistical research. It therefore publishes peer-reviewed articles representing original research in the broad areas of theoretical statistics, probability and applied statistics to pursue vigorous research activities. Thus it has played a decisive role to the advancement and dissemination of statistical information throughout the world, which is highly regarded by the peers.

In terms of visibility, articles in this journal are abstracted and reviewed in Mathematical Reviews (MR), Statistical Theory and Method Abstracts (STMA), Zentralblatt fur Mathematik and also indexed in Current Index to Statistics (CIS) and Scopus. In fact, Science Citation Index (SCI) used to cover Sankhyā for a period from 1966 to 1992 (as found in http://apps.webofknowledge.com/). Although, some causes on the coverage of Indian journals in SCI are explained in general, but no specific reason have been found for this journal (Satyanarayan and Jain 2002). However, Scopus covers this journal with SNIP (Source Normalized Impact per Paper) and SJR (ScImago Journal Rank) value of 0.032 and 0.072 respectively in 2011. Basically, SNIP corrects for differences in the frequency of citation across research fields, and SJR reflects the prestige of source as well as value the weighted citations per document. The journal publishes quarterly issues usually come out in February, May, August and November. Further details about the journal are available at http://sankhya.isical.ac.in.

Keeping in view of the aforesaid objectives, the primary data for the study has been collected from MathSciNet (2010). It enables web access to Mathematical Reviews (MR) database via multiple mirror sites and offers excellent content with powerful search functionality and timely updates. Dynamic search interface of MR provides diverse searchable fields including author affiliations, institution-code, country code, classification code, and source journal name that could be useful to identify the articles of particular journal across different time-frame. In fact, Boolean operators can effectively create many different combinations among the fields. Therefore, bibliographic data of the articles having source-journal as Sankhya in the byline and published during 2003 to 2007 were retrieved from the MathSciNet database. Complete searching displayed 199 hit records that are found a reasonable sample size for the purpose of this study. Prior to tabulation, retrieved data set is verified with the physical volumes of the journal available in the Indian Statistical Institute library collection. Ultimately, various bibliometric techniques are

applied to determine the patterns of publications and extent of collaborations, as well as geographical and institutional distribution of authors. Bibliographic data such as year of publication, author name, number of authors, affiliations, collaboration types, mathematics subject classification, and time lag of publication, are recorded and subsequently analyzed for making observations and interpretations.

Collaborative research has been assessed on the basis of the following quantitative indicators: proportion of non-collaborative (single author) and collaborative (those with two or more authors) papers; proportion of papers corresponding to different types of collaboration like institution-wise, country-wise; and increase of collaborative papers over the years. In addition, degree of collaboration by means of collaborative coefficient (CC) has been estimated using Subramanyan's formula (Subramanyam 1983). For productivity of authors, Lotka's law (Lotka 1926) is applied and tested. A ranked list of prolific authors is also prepared based on the weighted values of publications using adjusted or fractional counting method (Van-Hooydonk 1997). Distribution of articles across the sub-domains has been made based on the American Mathematical Society (AMS) primary classification code (three-digit level) as available in the source database. Thus, a thorough analysis of collected data has been worked out in different dimensions using various mathematical and statistical techniques. Finally, necessary data sheets are tabulated and illustrated for interpretation and drawing conclusions.

FINDINGS

Year-wise Distribution of Contributions

Table 1 presents the year-wise distribution of articles in the journal during the study period. A total of 199 papers were published during 2003 to 2007, distributed over 20 issues. An average of 9.95 articles is contributed to each issue of this journal. The number of contribution decreased consistently over the years except in 2007. Significantly the growth trend of publications in Sankhyā is found to be almost similar to the overall growth trend of statistical publications produced in India over the same period, as shown in Figure 1. Statistical publications of India (given under India-total) were obtained from the MathSciNet database using the search expression "*(institution code=(6-*) AND MSC primary=(62*) AND publication type=(Journals)) AND pubyear=2003*". In the search expression, 6- indicates the country-code for India, which follows a particular institution-code (eg. 6-ISI); whereas 62 denotes the subject-code for statistics, as assigned in the mathematics subject classification of AMS.

Table 1: Year-wise Distribution of Articles

Year	Volume (issue)	No of articles	Percentage	India-total	Percentage
2003	65 (1-4)	49	24.62	201	23.87
2004	66 (1-4)	45	22.61	184	21.85
2005	67 (1-4)	38	19.09	161	19.12
2006	68 (1-4)	28	14.07	142	16.86
2007	69 (1-4)	39	19.59	154	18.29
Total	Five (twenty)	199	100	842	100

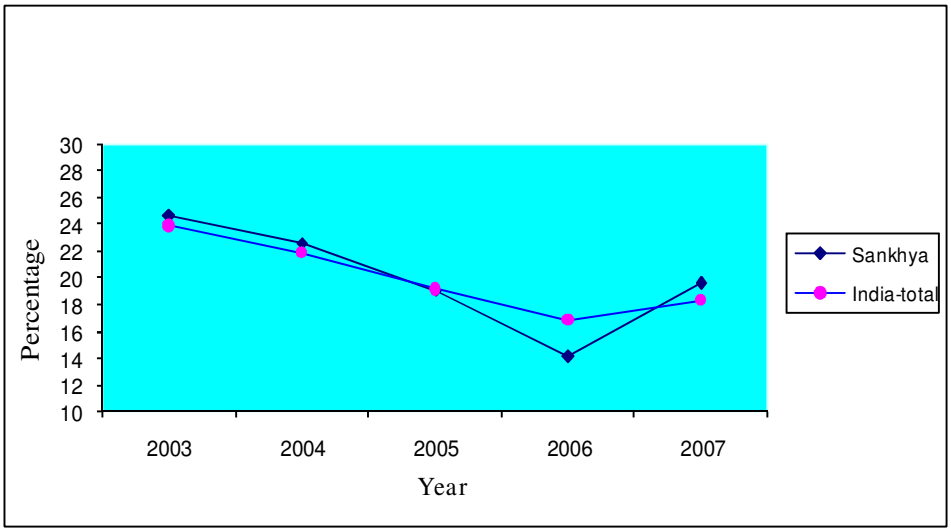


Figure 1: Growth Trend of Statistical Publications (Sankhya vs. India-total)

Authorship Pattern

Table 2 presents the authorship pattern observed in the contributions published in Sankhyā during 2003-2007. It shows a total of 392 occurrences of authors counted in 199 articles produced during the period, thus the average authorship obtained is 1.96 for each publication. It is observed that single-authored papers are quite significant (30.15%), although the majority of Sankhyā contributors worked in highly collaborative manner (69.85%). Articles produced by collaboration of two-authors (46.73%) are most predominant, which is followed by three authors (19.59%) and four authors (3.01%). However, an increasing trend of multi-authored publications (from 69% in 2003 to 72% in 2007) has been observed in agreement with many other disciplines. Bandyopadhyay (2001), in a study on authorship pattern in different disciplines observed that multiple-authorship trends have increased steadily through decades. Kalyane and Sen (1995) noted the increase of multi-authored publications in various fields such as agriculture, economics, psychology, life sciences and medicine. Visakhi and Srivastava (2002) in their study on research collaboration in statistical science also endorsed the similar pattern. Such a trend of cooperation among scientists is perhaps due to the increased complexity of research, technological expositions combined with more specialization, cost of modern investigations, and often interdisciplinary research areas have been forcing the researchers to share their expertise in contributing the articles.

Table 2: Distribution of Articles by Authorship

Year	Number of Articles	Authorship value					Occurrence of authors	Average authorship
		Single	**Two**	**Three**	**Four**	**Five**		
2003	49	15	27	6	1	0	91	1.857
2004	45	12	22	9	2	0	91	2.022
2005	38	12	16	8	2	0	76	2.000
2006	28	10	14	4	0	0	50	1.785
2007	39	11	14	12	1	1	84	2.153
Total	199	60	93	39	6	1	392	1.969

Ranking of Prolific Authors

Table 3 enumerates the ranking of prolific authors based on the weighted value of their contributions in Sankhyā during the period of study. Weighted value of contributed articles is calculated using fractional counting method; where the total weight of an article is always considered 1, where it assumes that each author contributed equally to a paper and adjusts for authorship (Abrizah and Wee 2011). This method can produce more accurate values in making the differences with finer tunes and removes anonymous ranking of authors as yielded from direct counting method (Egghe, Rousseau and Van-Hooydonk 2000). For instance, those authors actually ranked first, second, and third (having the weighted value 4.000, 2.333, 2.167 respectively) producing 4 articles each would come to the top, if direct counting method were applied.

Table 3: Ranked List of Prolific Authors based on Weighted Value of Contributions

Rank	Author name (affiliation code)	Authorship in contributions					Number of contribution	Weighed value
		Single	Two	Three	Four	Five		
1	Jacroux, Mike (1-WAS-S)	4	-	-	-	-	4	4.000
2	Sutradhar, Brajendra C. (3-NF)	1	2	1	-	-	4	2.333
3	Das, Ashish (6-ISIND)	1	1	2	-	-	4	2.167
4	Bhattacharjee, M. C. (1-NJIT-AM)	2	-	-	-	-	2	2.000
4	Cheng, Fuxia (1-ILS)	2	-	-	-	-	2	2.000
4	Pommeret, Denys (F-CREST-ENSAI)	2	-	-	-	-	2	2.000
5	Lahiri, Soumendra N. (1-IASU-S)	1	1	1	-	-	3	1.833
6	Chen, Pingyan (PRC-JNNU)	1	1	-	-	-	2	1.500
6	James, Lancelot F. (PRC-HKST-SMG)	1	1	-	-	-	2	1.500
6	Jurečková, Jana (CZ-KARL-S)	1	1	-	-	-	2	1.500
6	Kundu, Debasis (6-IITK)	1	1	-	-	-	2	1.500
6	Lin, Gwo Dong (RC-AST-S)	1	1	-	-	-	2	1.500
6	Meintanis, Simos G. (GR-UATH-EC)	1	1	-	-	-	2	1.500
7	Pensky, Marianna (1-CFL-S)	1	-	1	-	-	2	1.333
8	Berti, Patrizia (I-MORE-PM)	-	1	2	-	-	3	1.167
8	Rigo, Pietro (I-PAVI-PL)	-	1	2	-	-	3	1.167
9	Forty authors having – each	1	-	-	-	-	1	1.000
9	Seven authors having – each	-	2	-	-	-	2	1.000
10	Thirteen authors having – each	-	1	1	-	-	2	0.833
11	Two authors having – each	-	1	-	1	-	2	0.750
12	Mallick, Bani K. (1-TXAM-S)	-	1	-	-	1	2	0.700
13	Two authors having – each	-	-	2	-	-	2	0.667
14	Two authors having – each	-	-	1	1	-	2	0.583
15	Ghosh, Malay (1-FL-S)	-	-	1	-	1	2	0.533
16	One-forty-four authors having – each	-	1	-	-	-	1	0.500
16	Four authors having – each	-	-	-	2	-	2	0.500
17	Eighty-eight authors having – each	-	-	1	-	-	1	0.333
18	Twelve authors having – each	-	-	-	1	-	1	0.250
19	Three authors having – each	-	-	-	-	1	1	0.200
Total	335 unique authors	60	186	117	24	5	392	199

Table 3 shows a total of 336 unique authors having 392 occurrences in different authorship positions of 199 contributions. It is observed that 16 individual authors having the weighted value >1 for their contributions occupied the top eight ranks. Mike Jacroux

(Washington State University, USA) is found to be most prolific author followed by Brajendra C. Sutradhar (Memorial University of New Foundland, Canada); Ashish Das (Indian Statistical Institute, New Delhi); M. C. Bhattacharjee (New Jersey Institute of Technology, USA); Fuxia Cheng (Illinois State University, USA); Denys Pommeret (CREST-ENSAI, France); and Soumendra N. Lahiri (Iowa State University, USA). In the ranked list, the name of the contributors who received weighted score ≤ 1 is not revealed.

Applicability of Lotka's Law

Lotka's empirical law of scientific productivity states that y number of authors each credited with x number of papers is inversely proportional to x, which is the output of each individual author. Thus relation is expressed as (Lotka 1926):

$$x^n \alpha \frac{1}{y} \quad \text{or} \quad x^n y = C \quad \ldots\ldots\ldots\ldots \text{(i)} \quad [\text{n and C are two constants}]$$

There has been a considerable literature on the empirical validation of Lotka's law. Several studies have reported that Lotka's law is applicable for the productivity trend distributions of well-recognized disciplines. Usually such disciplines follow the distribution patterns that conform to Lotka's law in its original form with exponent value of 2. While some other investigations found that the value of exponent n is not always 2, rather a variable value around 2.

Murphy (1973) applied the Lotka's law appropriately in the field of humanities, without any statistical test to check the degree of significance. Miranda Lee Pao presented the step by step application process of Lotka´s law, deducing the values of the constant and the exponent based on the method of Lotka, as well as tested the degree of significance (Pao 1985). Later she applied this procedure over 48 groups of authors (representing 20 scientific disciplines) and found that in most of the cases the original law of Lotka holds good (Pao 1986). Nicholls (1986) conducted studies on 15 different datasets of humanities, social sciences, and sciences for testing the empirical validation of the law. He observed that the studies on their majority are conflicting, incomparable, and inconclusive; thus do not provide any clear-cut validation of the Lotka's law. Such inconsistencies in validation of the law are perhaps due to a steady increase of co-authored publications over the time. Potter (1981) discussed in a review that Lotka credited only the senior author for each contribution ignoring all co-authors, as multi-authorship contribution was less common during Lotka's time. However, a number of studies showed that using total or even fractional counting of authorship lead to a breakdown of Lotka's law (Rousseau 1992).

Therefore, instead of commonly used inverse square law, Lotka's formulation can be observed as inverse power law in general, i.e. $x^n \cdot y = C$. The exponent (n) and the constant (C) can be estimated from the given set of author productivity data. A generalized form of Lotka's law (referred to inverse power law) as presented by Bookstein (1976) could be useful here:

$$a_n = \frac{C}{n^\alpha} \quad \text{for } n = 1, 2, 3 \ldots \text{ and } C > 0 \quad \ldots\ldots\ldots\ldots \text{(ii)}$$

Where a_n represents the probability of authors producing n contributions each and C and α are two parameters to be estimated for a specific set of data. The value of productivity constant (α) or characteristic exponent can be determined by considering the values of n (1, 2, 3…) applying either graphical or mathematical method. Here an attempt has been made to predict simply on the applicability of Lotka's law for author productivity in the

sample dataset; and to what extent author's productivity conforms to Lotka's law has also been carried out. Table 4 shows the author productivity considering all the authors; where 287 authors have one paper each, 42 authors produced only two papers each, 3 authors contributed three papers each, and another 3 authors have four papers each to their credit. The maximum number of papers that have been credited to an individual author is found to be four only. Now considering the observed data (i.e. 287 authors produced 1 paper each), anyone can easily derive the value of C from the equation (ii).

$$a_n = \frac{C}{n^{\alpha}} \quad \text{or,} \quad 287 = \frac{C}{1^{\alpha}} \quad \text{or,} \quad C = 287$$

Subsequently, taking the expected value of α as 2 and putting the derived value of C as well as values of n (1, 2, 3, 4) in the above equation, corresponding values of expected authors (a_n) are obtained. Result shows (Table 4) a considerable variation in the expected values when compare to observed values. So, the law does not fit in this case and a violation is clearly observed.

Table 4: Author productivity in Sankhyā during 2003-2007 (All authors considered)

No of articles (A)	No. of authors Observed (B)	Percentage (%)	Authorship (A x B)	Percentage (%)	No. of authors expected when α = 2	No. of authors expected when α = 2.77
1	287	85.671	287	73.214	287	287
2	42	12.537	84	21.428	72	42
3	3	0.895	9	2.295	32	13
4	3	0.895	12	3.061	18	6
Total	335	100	392	100	409	348

It is also evident from Table 4 that when the value of α (productivity parameter) approximated to 2.77 (instead of 2) then the expected values of a_n are quite close to the observed values.

$$a_n = \frac{C}{n^{\alpha}} \quad \underline{\text{or}} \quad n^{\alpha} = \frac{C}{a_n} \quad \underline{\text{or}} \quad \log n^{\alpha} = \log \frac{C}{a_n} \quad \underline{\text{or}} \quad \alpha \log n = \log \frac{C}{a_n}$$

$$\underline{\text{or}} \quad \alpha = \frac{\log \frac{C}{a_n}}{\log n} \quad \underline{\text{or}} \quad \alpha = \frac{\log \frac{287}{42}}{\log 2} \quad [\textit{for } C = 287,\ a_n = 42,\ n = 2]$$

$$\underline{\text{or}} \quad \alpha = \frac{0.83433}{0.30103} = 2.77$$

Putting the values of n (1, 2, 3, 4) and calculated value of α as 2.77 the following values of a_n are derived:

$$a_n = \frac{C}{n^{\alpha}} = \frac{287}{1^{2.77}} = 287\ ;\ a_n = \frac{287}{2^{2.77}} = 42.07\ ;\ a_n = \frac{287}{3^{2.77}} = 13.68\ ;\ a_n = \frac{287}{4^{2.77}} = 6.16$$

Table 4 depicts that productivity distribution data partially fits the Lotka's law in its original form with a calculated value of exponent α = 2.77 and the number of papers does not exceed two. The law does not hold well beyond this value. It is worthy to mention that the larger the value of α is, the greater is the gap between the productivity of individual groups

of authors contributing n number of papers each. Practically a larger value of α implies the proportion of highly productive authors is decreased (Gupta 1995). Further statistical tests (such as chi-square of goodness-of-fit and K-S test) could be useful to confirm the applicability of this law at the appropriate levels of significance.

Degree of Collaboration

Research collaboration is very much common in any scientific field and is highly practised in the twenty-first century. Collaboration is an intense form of interaction that allows for effective communication as well as sharing of competence and other resources. However, multiple-authorship in different dimension (such as inter-institution and inter-country) provides a measure of intensity in collaborations. Table 5 reveals the collaboration scenario among the authors in different three levels – namely Indian (within authors from India), international (authors within a country other than India), and combined (authors from two or more different countries). Out of 139 multi-authored contributions, Indian collaboration constitutes only 8% and international collaboration constitutes 58%, while the share of 34% of multi-authored contributions is collaborated among statisticians across the countries. Clearly it brings out the prevalence of collaborative research (69.85%) over the single research (30.15%) in the contributions of Sankhyā.

Table 5: Collaboration Trend among Authors and Collaborative Coefficient.

Year	Non-collaborative (N_s)	%	Collaborative (N_m)				%	CC
			Indian	Inter-national	Combined	Total		
2003	15	30.62	3	17	14	34	69.38	0.6938
2004	12	26.67	3	19	11	33	73.33	0.7333
2005	12	31.58	2	13	11	26	68.42	0.6842
2006	10	35.71	2	9	7	18	64.29	0.6429
2007	11	28.21	1	23	4	28	71.79	0.7179
Total	60	30.15	11	81	47	139	69.85	0.698

In order to measure the degree of collaboration in quantitative terms, the formula given by Subramanyam (1983) can be useful. Subramanyam worked out the collaborative coefficient (CC), which is determined by the ratio of number of collaborative publications and total number of publications during certain period of time. That can be expressed as:

$$CC = \frac{N_m}{N_m + N_s} = \frac{139}{139 + 60} = 0.698$$

Where N_m refers to multi-authored (two or more) contributions and N_s denotes the number of single-authored contributions published in the journal during the study period. Thus, the average degree of collaboration is found to be 0.69 and is quite significant. The extent of collaboration distribution over the period is presented in Figure 2. Clearly it indicates the prevalence of team or group research in the field studied here, i.e. scientists working in the field of statistics prefer to conduct research in collaboration. Similar observation has also been found in a study on research trends in the field of statistical science (Visakhi and Srivastava 2002).

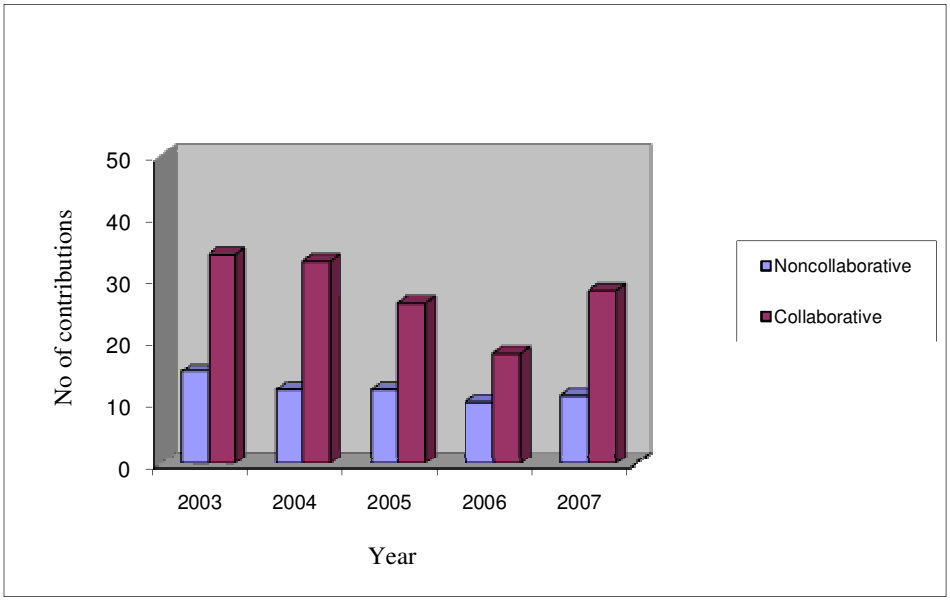


Figure 2: Collaboration Trend over the Years (Single vs. Multi-authors)

Bilateral and Multilateral Collaboration Density

The lateral relationship among co-authors of collaborative contributions can be studied under three different levels of aggregation – unilateral, bilateral, and multilateral. Unilateral collaboration is described when co-authorship of a publication occurs within a link, whereas bilateral collaboration implies the co-authorship occurs between two different links. Multilateral collaboration indicates the participation of co-authors from more than two different links for producing an article. Table 6 depicts the distribution of collaborative contributions in order to map the lateral relationship among co-authors.

Table 6: Lateral Relations among Collaborative Contributions

Year	Non-collaborative	Institution wise collaboration				Country wise collaboration			
		Uni-lateral	Bi-lateral	Multi-lateral	Total	Uni-lateral	Bi-lateral	Multi-lateral	Total
2003	15	4	28	2	34	19	14	1	34
2004	12	13	14	6	33	21	11	1	33
2005	12	5	16	5	26	15	9	2	26
2006	10	4	12	2	18	11	6	1	18
2007	11	9	13	6	28	24	3	1	28
Total	60	35	83	21	139	90	43	6	139

Collaborative contributions are viewed in two different angles, i.e. institute-wise collaboration and country-wise collaboration. Institute-wise collaboration of a publication happens to be made by the authors; either from the same institution (unilateral), or from two different institutes (bilateral), or may be from more than two different institutions. Similarly, country-wise multilateral collaboration of a publication implies that the author's affiliated institutions are located in three or more different countries. For example, four authors have contributed to a publication; one is affiliated to University of California (USA), another one from the Indian Statistical Institute (India), and the other two authors are

affiliated to Iowa State University (USA) – can be a case of inter-institution-multilateral as well as inter-country-bilateral collaboration. Such indicator identifies the intent of collaborative research and helps to determine the strength of a research network. A considerable number of bilateral and multilateral collaboration (both inter-institution and inter-country) signifies that intellectual perceptions of diverse origin have been intermingled into this communication channel.

Country-wise Distribution of Authors

Table 7 shows the geographical distribution of contributing authors in Sankhyā during the study period. Country names have been identified from their affiliations as reflected in their respective publications, primarily available from the 'institution code' data-field of MathSciNet, subsequently verified in the physical volumes of the source journal. Tabulated data shows that a total of 392 contributors from 43 countries took part in producing 199 articles in Sankhyā during the study period. The number-of-countries represented for publishing papers, can be considered as the simplest indicator to measure the internationality of a journal (Perneger and Hudelson 2007). So the status of internationality of the source journal is clearly observed, i.e. the journal considerably gained diverse experiences and opinions in publishing the articles.

Table 7: Geographical Diversity of Contributing Authors.

Country Name	Country code	Number of Occurrence	Share value of contributions					Weighted value
			Full	1/2	1/3	¼	1/5	
United States of America	1	124	22	66	27	4	5	66.00
India	6	36	3	23	10	0	0	17.83
Canada	3	31	2	13	10	6	0	13.33
Italy	I	28	2	14	12	0	0	13.00
France	F	18	7	4	7	0	0	11.33
Germany	D	16	1	12	3	0	0	8.00
Peoples Rep. of China	PRC	12	5	4	1	2	0	7.83
Taiwan (R.O.C)	RC	18	2	3	9	4	0	7.50
Greece	GR	13	2	4	7	0	0	6.33
United Kingdom	4	13	1	5	7	0	0	5.83
Japan	J	8	3	2	3	0	0	5.00
Czech Republic	CZ	9	1	4	3	1	0	4.25
Israel	IL	5	2	2	1	0	0	3.33
Brazil	BR	7	0	2	3	2	0	2.50
Switzerland	CH	4	1	3	0	0	0	2.50
Spain	E	4	0	4	0	0	0	2.00
Finland	FIN	5	0	1	0	4	0	1.50
Belgium	B	4	0	1	3	0	0	1.50
Iran	IR	3	0	3	0	0	0	1.50
2 countries (Australia & Sweden)	5 & S	2 each	1	1	0	0	0	1.50 (x 2)
Chile	RCH	3	0	1	1	1	0	1.08
3 countries (Republic of Korea, Morocco, & Portugal)	KR, MRC, P	2 each	0	2	0	0	0	1.00 (x 3)
4 countries (United Arab Emirates, Nigeria, Mexico & Kenya)	UAE, WAN, MEX, KEN	1 each	1	0	0	0	0	1.00 (x 4)
Singapore	SGP	2	0	1	1	0	0	0.83
2 countries (Saudi Arabia &Turkey)	SAR, TR	2 each	0	0	2	0	0	0.66 (x 2)
6 countries (Russia, Poland, Oman, Jordon, Estonia & Cyprus)	RS, PL, OM, JOR, ES, CY	1 each	0	1	0	0	0	0.50 (x 6)
5 countries (The Netherlands, Malaysia, Luxembourg, Hungary & Algeria)	NL, MAL, LUX, H, DZ	1 each	0	0	1	0	0	0.33 (x 5)
Total 43 countries		392	60	186	117	24	5	199

A ranked list of participating countries was prepared on the basis of weighted share value of the contributions from respective countries, applying fractional counting method. USA receives the maximum weight of 66 (33%) by affiliating 124 occurrences of authors with different authorship values; followed by India (9%), Canada (7%), Italy (6.5%), France (6%), and Germany (4%). It has been found that top ten countries are carrying about 80% of the total weight and each of them having greater than 5.0 weighted value. The rest of the weight (20%) is eventually distributed over 33 countries. This indicator helps to identify the partner countries having similar research interests and the extent of their involvement in recognizing the international repute of the journal as well.

Institution-wise Distribution of Authors

Table 8 depicts the distribution of authors from various institutions who made their contribution to Sankhyā. The distributed data shows a total of 392 contributors from 309 individual institutions were involved in generating 199 papers in Sankhyā during the study period. A ranked list of participating institutions was prepared based on the weighted value of the contributions from respective institutions. Weighted value has been calculated considering proportionate representation of authorship in contributions produced by a particular institution. It has resulted more distinct list for determining the ranks of the institutes. It is observed from Table 8 that the Indian Statistical Institute, New Delhi (6-ISIND) appeared on the top; followed by Universita di Pavia, Italy (I-PAVI). Both the institutes contributed equal number of papers (12 each), however they ranked differently due to the unequal share value (6.00 & 5.33) of their contributions. Active participation of various institutions across geographical boundaries implies the recognition and authoritativeness of this journal in statistical research, as evident from the list.

Table 8: Institute-wise Distribution of Contributing Authors

Institute Name	Institute code	No. of Occurrence	Share value of contributions					Weighted value
			Full	1/2	1/3	1/4	1/5	
Indian Statistical Institute, ND	6-ISIND	12	1	8	3	0	0	6.00
Universita di Pavia, Italy	I-PAVI	12	1	4	7	0	0	5.33
Washington State University, USA	1-WAS	6	4	2	0	0	0	5.00
Indian Statistical Institute, Kolkata	6-ISI	8	1	7	0	0	0	4.50
Memorial University of New Foundland, Canada	3-NF	8	1	5	2	0	0	4.17
University of Athens, Greece	GR-UATH	8	1	4	3	0	0	4.00
University of California, USA	1-UCLA	7	1	5	1	0	0	3.83
Iowa State University, USA	1-IASU	5	2	1	2	0	0	3.17
University of Minnesota Twin Cities, USA	1-MN	4	2	2		0	0	3.00
Texas A & M University, USA	1-TXAM	6	1	2	1	0	2	2.73
University of Connecticut, USA	1-CT2	4	1	2	1	0	0	2.33
4 Institutes having CREST-ENSAI (France), Illinois State University (USA), New Jersey Institute of Technology (USA), Purdue University (USA).	F-ENPC-CRE, 1-ILS, 1-NJIT, 1-PURD	2 each	2	0	0	0	0	2.00 (x 4)
Karlovy (Charles) University (UK), Czech Republic	CZ-KARL	3	1	2	0	0	0	2.00
Ruhr-Universitat Bochum, Germany	D-BCHMM	4	0	4	0	0	0	2.00

3 Institutes having Academia Sinica, Institute of Statistics (Taiwan, R.O.C), Indian Institute of Technology - Kanpur, Tel Aviv University (Israel)	RC-AST, 6-IITK, IL-TLAV	3 each	1	1	1	0	0	1.83 (x 3)
4 Institutes having Pennsylvania State University (USA), Universita Commerciale - Luigi Bocconi (Italy), University of Manchester (UK), University of South Florida (USA)	1-PAS, I-UCOM, 4-MANC, 1-SFL	4 each	0	3	1	0	0	1.83 (x 4)
University of Central Florida	1-CFL	3	1	0	2	0	0	1.67
University of Michigan	1-MI	3	1	0	1	0	1	1.53
University of Missouri	1-MO	4	0	2	1	0	1	1.53
6 Institutes having Carleton University (Canada), Hong Kong University of Science and Technology (P.R. China), Jinan University (P.R. China), University of Kentucky (USA), University of Texas Pan American (USA), University of Tokyo (Japan)	3-CARL, PRC-HKST, PRC-JNNU, 1-KY, 1-PAM, J-TOKYOEC	2 each	1	1	0	0	0	1.50 (x 6)
National Sun Yat-Sen University (Taiwan, R.O.C)	RC-SYS	4	0	1	3	0	0	1.50
2 Institutes having McGill University (Canada), University of Wisconsin(USA)	3-MGL, 1-WI	5 each	0	1	0	4	0	1.50 (x 2)
National Chiao Tung University (Taiwan, R.O.C)	RC-NCT-S	5	0	0	2	3	0	1.42
Sichuan University (P.R. of China)	PRC-SUN	2	1	0	1	0	0	1.33
2 Institutes having Massachusetts Institute of Technology (USA), University of Manitoba (Canada)	1-MIT, 3-MB	3 each	0	2	1	0	0	1.33 (x 2)
Indian Council of Agricultural Research (ICAR)	6-ICAR-I	4	0	0	4	0	0	1.33
University of Waterloo (Canada)	3-WTRL-S	4	0	0	3	1	0	1.25
3 institutes having Universita di Modena e Reggio Emilia (Italy), Universite de Toulouse III-Paul Sabatier(France), University Catholique de Louvain (Belgium)	I-MORE, F-TOUL3, B-UCL	3 each	0	1	2	0	0	1.17 (x 3)
University of Florida (USA)	1-FL	3	0	1	1	0	1	1.03
23 institutes having	-	1 each	1	0	0	0	0	1.0 (x 23)
18 institutes having	-	2 each	0	2	0	0	0	1.0 (x 18)
2 institutes having Georg-August-Universitat zu Gottingen (Germany), Masaryk University (Czech Republic)	D-GTN-ST, CZ-MASS	3 each	0	0	3	0	0	1.0 (x 2)
University of Tampere (Finland)	FIN-TAM	4	0	0	0	4	0	1.00
4 institutes having	-	2 each	0	1	1	0	0	0.83 (x 4)
University of Sao Paulo – USP (Brazil)	BR-SPL	3	0	0	1	2	0	0.83
Beijing (Peking) University (P.R.China)	PRC-BJ	2	0	1	0	1	0	0.75
7 institutes having	-	2 each	0	0	2	0	0	0.67 (x 7)
Catholic University of Chile	RCH-UCC	2	0	0	1	1	0	0.58
63 institutes having	-	1 each	0	1	0	0	0	0.5 (x 63)
38 institutes having	-	1 each	0	0	1	0	0	0.33(x38)
4 institutes having	-	1 each	0	0	0	1	0	0.25 (x 4)
209 individual institutions		392	60	186	117	24	5	199

Subject-wise Distribution of Articles

One of the objectives of this study is to ascertain the subject clusters that are predominating in this scholarly communication channel of statistical research. In view of this objective, subject areas pertaining to the articles are identified based on the primary subject code (MSC Code of AMS) in two-digit level, as assigned in each articles. Distribution of articles in broad subject clusters is presented in Table 9. Evidently two subject domains (Statistics; Probability theory and stochastic processes) cover almost 96% of the contributed articles. The rest of the subject domains constituted only 4% of the articles.

Table 9: Distribution of Articles in various Subject Clusters

Domain name	MSC Code	Frequency	Percentage
Statistics	62	153	76.884
Probability theory and stochastic processes	60	38	19.095
Operations research and mathematical programming	90	4	2.010
Biology and other natural sciences	92	2	1.005
Statistical mechanics and structure of matter	82	1	0.503
Measure and integration	28	1	0.503
Total		199	100

Further distribution of 153 articles belonging in Statistics (62-) has been analysed to identify the active sub-domains of this subject cluster. Subdivisions have been determined by the AMS subject classification code in three-digit level. Sub-domain wise distribution shows that the authors have pursued their research mostly in the areas of nonparametric inference (62-G), parametric inference (62-F), and followed by design of experiments (62-K), multivariate analysis (62-H), linear inference (62-J), statistical distribution theory (62-E), and decision theory (62-C). All other sub-domains (excluding statistics) together covered less than 12% of the total articles, as displayed in Figure 3.

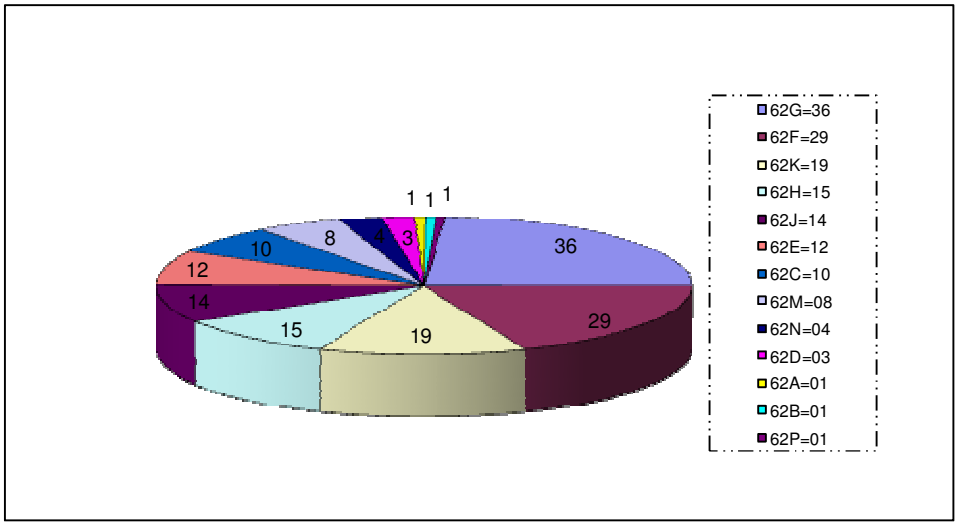


Figure 3: Distribution of Sub-domains of Statistics

Time Lag in Publications

Scientific journals are often criticized for the time lag in publishing the manuscripts. Time lag refers to the time taken between the date of receipt of a manuscript and its publication in the journal. A lengthy time-gap may affect on the intended impacts of intellectual outputs produced by the researchers. In the case of *Sankhyā*, time lags (in months) of all articles have been counted separately and grouped into various time-slots. Subsequently, the number of articles (frequency) belonged to a group is tabulated and the statistical mean value of time lag has been calculated, as shown in Table 10. An optimum time lag of fifteen months (Mean = 14.85) is found in the period of study.

Table 10: Frequency Distribution for Calculating the Mean Value of Time Lag

Time-gap	Frequency (f)	Mid value (m)	f x m	Mean value
01 to 05 months	21	3	63	$\overline{X} = \frac{\sum fm}{N}$ $= \frac{2955}{199}$ **=14.85**
06 to 10 months	54	8	432	
11 to 15 months	49	13	637	
16 to 20 months	37	18	666	
21 to 25 months	15	23	345	
26 to 30 months	8	28	224	
31 to 35 months	6	33	198	
36 to 40 months	5	38	190	
41 to 49 months	2	45	90	
50 to 60 months	2	55	110	
Total	199 (N)		2955	

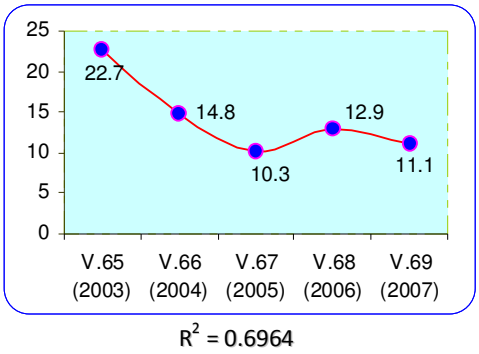
The average time lag of individual issues and volumes are also determined sequentially. Table 11 and the adjacent graph (Figure 4) depict the average time lag through issues and volumes of Sankhyā during the period of study. It is astonishing to note that average-time-lag over the issues has found to be maximum 32 months (in Part-1 of 2003) and minimum 9 months (Part-3 of 2007) for publishing the manuscripts in this scholarly journal. Volume wise average-time-lag was also higher in 2003. In fact, the calculated time lag makes a substantial gap, when added to a common delay exists in the journal. It seems very unrealistic that the revision dates (for example December 2007) of a few articles have superseded the publication date (May 2007) of the submitted manuscripts (May 2007) – for instance, the paper published in *Sankhyā, 2007, Vol.69, no.2, p.372*. It is possible when a journal either suffers from delays or suppresses the delays in publishing the issues more than a calendar year.

The time lag in the case of Sankhyā has no consistency for the articles published during study period, although a tendency of decreasing time lag over the years has been observed, as presented in Figure 4. The graph shows a declining trend of time lag following second-degree polynomial type (R^2 = 0.6964). The lower the lag is has always been encouraging for authors or researchers to pursue their research communications in any peer-reviewed journal, and the editorial board should pay much attention to minimize the time lag in future publications of a particular journal.

Table 11: Average Time Lag over the Issues and Volumes (in months)

Vol. (issue), Date	No. of articles	Average lag/ issue	Average lag/ vol.
65 Pt-1, Feb. 2003	12	31.8	22.7
65 Pt-2, May 2003	15	21.7	
65 Pt-3, Aug. 2003	13	13.7	
65 Pt-4, Nov. 2003	9	23.7	
66 Pt-1, Feb. 2004	10	18.1	14.8
66 Pt-2, May 2004	12	15.1	
66 Pt-3, Aug. 2004	12	14.3	
66 Pt-4, Nov. 2004	11	11.5	
67 Pt-1, Feb. 2005	6	10.7	10.3
67 Pt-2, May 2005	13	9.6	
67 Pt-3, Aug. 2005	8	9.8	
67 Pt-4, Nov. 2005	11	11.0	
68 Pt-1, Feb. 2006	7	10.1	12.9
68 Pt-2, May 2006	8	14.1	
68 Pt-3, Aug. 2006	6	11.0	
68 Pt-4, Nov. 2006	7	16.1	
69 Pt-1, Feb. 2007	6	10.5	11.1
69 Pt-2, May 2007	12	10.2	
69 Pt-3, Aug. 2007	9	9.1	
69 Pt-4, Nov. 2007	12	14.6	

Figure 4: A Graph Showing Volume Wise Time Lag



CONCLUSION

Scientometric measurements have been recognized as an indispensable tool for intelligent judgment of research activities and scientific behaviours. The analyses presented in this study have permitted many conclusions of broad generality on statistical science research and in particular to Sankhyā. We find that the growth trend of publications in Sankhyā conforms to the overall growth trend of statistical publications in India that are produced during the study period. Like many other disciplines, an increasing trend of collaboration among the researchers has been observed in this scientific field, where bilateral and multilateral collaboration (across the institutions and countries) is found quite significant. Prevalence in research-collaboration is probably due to the increased complexity of research (having technological expositions combined with more specialization) which often force the researchers to share their expertise; thus signifies that intellectual perception of the researchers from diverse origin has been intermingled into this scholarly Journal. It has been found that author productivity data is not in agreement with the Lotka's law in its' original form. Considerable variations in expected values when compared to observed values are noticed. Sankhyā, being one of the reputable journals in statistics, has maintained a careful balance in the frontier areas of research in Statistics and Probability theory, as committed in the editorial policy of the journal. These two subject clusters are steadily predominating in this journal and covered almost 96% of the total contributed articles during the study period. However, the sub-domains such as parametric and nonparametric inference, design of experiments, multivariate analysis, linear inference,

distribution theory, probability theory and stochastic processes have also become very active areas of research in this communication channel.

Findings also indicate that the journal suffers from lengthy time lags in publishing the manuscripts, which may affect on intended impacts of the intellectual outputs produced by the authors. Such delay often leads to poor citation, low impact factor, and even discourages the researchers from submitting their manuscript to this journal. Although this study has identified a decreasing trend of time lag, the editorial board should pay much attention to reduce it further. More dedicated editorial policies toward the timeliness, and web accessibility of intellectual contents through reputed publisher's platform would be a great deal in this regard. Certainly the Sankhyā bears an international appeal in terms of popularity (circulation & web-log hits), reputation (coverage in leading indexing and abstracting services), acceptance (diversity of author's participation), authoritativeness (depth of research), and above all its life span (exists since 1933), although these indicators might differ considerably depending on the individual researcher's perception of the journal (Theoharakis and Skordia 2003). It is indeed essential to look inward and seek the reasons why such a reputable scholarly journal is currently out of the coverage in the Social Science Citation Index (SSCI) of the Web of Science (WoS). So, there is need to understand and formulate strategies for the re-inclusion of Sankhyā within the purview of SSCI and WoS. Further analysis of Sankhyā may be performed to answer various questions emerged from the analyses of this study. The analysis of citation patterns, tracking of collaboration-network, and assessment of internationality of Sankhyā could be the probable areas of research in this direction. Simpson (1949), Shannon (1948), and Kwoka (1977) indices can be applied to the country-wise distribution data for measuring the diversity of contributed papers, which is an indicator to visualize the international appeal in Sankhyā.

REFERENCES

Abrizah, A. and Wee, M.C. 2011. Malaysia's Computer Science research productivity based on publications in the Web of Science, 2000-2010. *Malaysian Journal of Library & Information Science.* Vol. 16, no. 1: 109-124

Anyi, K. W. U., Zainab, A. N. and Anuar N B. 2009. Bibliometric studies on single journals: a review. *Malaysian Journal of Library & Information Science*, Vol. 14, no. 1: 17-55.

Bandyopadhyay, A. K. 2001. Authorship pattern in different disciplines. *Annals of Library and Information Studies*. Vol. 48, no. 4: 139-147.

Bookstein, A. 1976. The bibliometric distribution. *Library Quarterly*. Vol. 46, no.4: 416-423.

Egghe, L., Rousseau, R. and Van-Hooydonk, G. 2000. Methods for accrediting publications to authors or countries: consequences for evaluation studies. *Journal of the American Society for Information Science.* Vol. 52, no. 2: 145-157.

Gupta, D. K. 1995. Authorship trend and application of Lotka's law: early literature of computer based storage and retrieval of geoscientific data and information. *IASLIC Bulletin.* Vol. 30, no.1: 13-22.

Kalyane, V. L. and Sen, B. K.. 1995. A bibliometric study of the journal of Oilseeds research. *Annals of Library Science and Documentation*. Vol 4, no. 4: 121 – 141.

Kwoka, J.E. 1977. Large firm dominance and price-cost margins in manufacturing industries. *Southern Economic Journal*. Vol.44, no.1: 183-189.

Lotka, A. J. 1926. The frequency distribution of scientific productivity, *Journal of the Washington Academy of Science*, Vol.16: 317-323.

MathSciNet, American Mathematical Society (AMS), Providence, USA. Available at: http://www.ams.org/mathscinet.

Murphy, L. J. 1973. Lotka's Law in the Humanities. *Journal of the American Society for Information Science.* Vol. 24, no.6: 461-462.

Nicholls, P. T. 1986. Empirical validation of Lotka's law. *Information Processing and Management*. Vol. 22, no. 5: 417-419.

Pao, M. L. 1985. Lotka's law: a testing procedure. *Information Processing and Management*. Vol. 21, no. 4: 305-320.

Pao, M. L. 1986. An empirical examination of Lotka's law. *Journal of the American Society for Information Science.* Vol. 37, no. 1: 26-33.

Perneger, T. V. and Hudelson, P. M. 2007. How international is the International Journal for Quality in Health Care?, *International Journal for Quality in Health Care*. Vol. 19, no. 6: 329-333.

Potter, W. G. 1981. Lotka's law revisited. *Library Trends.* Vol. 30, no.1: 21-39.

Rousseau, R. 1992. Breakdown of the robustness property of Lotka's law: the case of adjusted counts for multiauthorship attributions. *Journal of the American Society for Information Science.* Vol. 43, no. 10: 645-647.

Rudra A. 1996. *Prasanta Chandra Mahalanobis: a biography*. Delhi: Oxford University Press, p.268-270.

Satyanarayana, K. and Jain, N. C. 2002. Coverage of Indian S&T journals in SCI/JCR. *Information Today and Tomorrow*. Vol. 21, no. 2: 6-11.

Shannon C.E. 1948. A mathematical theory of communication. *Bell System Technical Journal*. Vol.27, no.3: 379-423, 623–656.

Simpson, E.H. 1949. Measurement of diversity. *Nature*. Vol.163: 688.

Subramanyam, K.. 1983. Bibliometric studies of research in collaboration: a review. *Journal of Information Science.* Vol. 6, no. 1: 33-38.

Theoharakis, V. and Skordia, M. 2003. How do statisticians perceive statistics journals? *The American Statistician.* Vol. 57, no. 2: 115-123.

Van-Hooydonk, G. 1997. Fractional counting of multiauthored publications: consequences for the impact of authors. *Journal of the American Society for Information Science.* Vol. 48, no. 10: 944-945.

Visakhi, P. and Srivastava, S. S. 2002. Current trend of research collaboration in the field of statistical science - a case study. *IASLIC Bulletin.* Vol. 47, no. 4: 210-215.